

Ejemplo de Regresión Logística Multiclase con Descenso por Gradiente

Cristian Omar Alvarado Rodríguez - Machine Learning

22 de septiembre de 2025

1. Introducción

En este ejercicio se construyó un modelo de **regresión logística multiclase** para clasificar usuarios según el sistema operativo utilizado al visitar un sitio web.

Las clases fueron definidas como:

- 0 – Windows
- 1 – Macintosh
- 2 – Linux

Las características de entrada fueron:

1. Duración de la visita en segundos
2. Número de páginas vistas
3. Cantidad de acciones realizadas
4. Valor total de las acciones

El conjunto de datos estuvo compuesto por 170 registros, lo que permite realizar un ejemplo sencillo y comprensible, aunque en la práctica se recomienda utilizar un número mayor de muestras.

2. Modelo de Regresión Logística

Dado un vector de características de entrada $\mathbf{x} \in \mathbb{R}^n$ y los parámetros del modelo (pesos \mathbf{w}_k y sesgo b_k), la probabilidad de pertenecer a la clase k se calcula mediante la función **softmax**:

$$P(y = k \mid \mathbf{x}; \mathbf{W}, \mathbf{b}) = \frac{e^{\mathbf{w}_k^T \mathbf{x} + b_k}}{\sum_{j=1}^K e^{\mathbf{w}_j^T \mathbf{x} + b_j}}$$

donde K es el número total de clases.

3. Función de Pérdida

Para medir el error del modelo utilizamos la **entropía cruzada**, definida como:

$$J(\mathbf{W}, \mathbf{b}) = -\frac{1}{m} \sum_{i=1}^m \sum_{k=1}^K y_{i,k} \log P(y = k \mid \mathbf{x}^{(i)}; \mathbf{W}, \mathbf{b})$$

donde m es el número de ejemplos en el conjunto de entrenamiento y $y_{i,k}$ es igual a 1 si el ejemplo i pertenece a la clase k y 0 en caso contrario.

4. Descenso por Gradiente

El objetivo es minimizar la función de costo $J(\mathbf{W}, \mathbf{b})$. Para ello, se actualizan los parámetros del modelo mediante la siguiente regla:

$$\theta := \theta - \alpha \frac{\partial J}{\partial \theta}$$

donde:

- θ representa un parámetro del modelo (peso o sesgo),
- α es la tasa de aprendizaje,
- $\frac{\partial J}{\partial \theta}$ es el gradiente de la función de costo respecto al parámetro.

Este procedimiento se repite de forma iterativa hasta que la función de pérdida converge.

5. Resultados

Durante el entrenamiento se obtuvo la siguiente información:

- La curva de pérdida mostró cómo el error disminuye a lo largo de las iteraciones, evidenciando el aprendizaje del modelo.
- La matriz de confusión permitió evaluar la calidad de la clasificación entre Windows, Macintosh y Linux.

En la Figura 1 se observa la curva de pérdida, mientras que en la Figura 2 se muestra la matriz de confusión.

6. Conclusiones

Se comprobó que la **regresión logística multiclase**, optimizada con **descenso por gradiente**, es un modelo eficaz para problemas de clasificación discreta.

Aunque se trabajó con un conjunto reducido de datos (170 registros), se observó un proceso de aprendizaje adecuado. En la práctica, contar con un mayor número de muestras proporcionaría un modelo más robusto y generalizable.

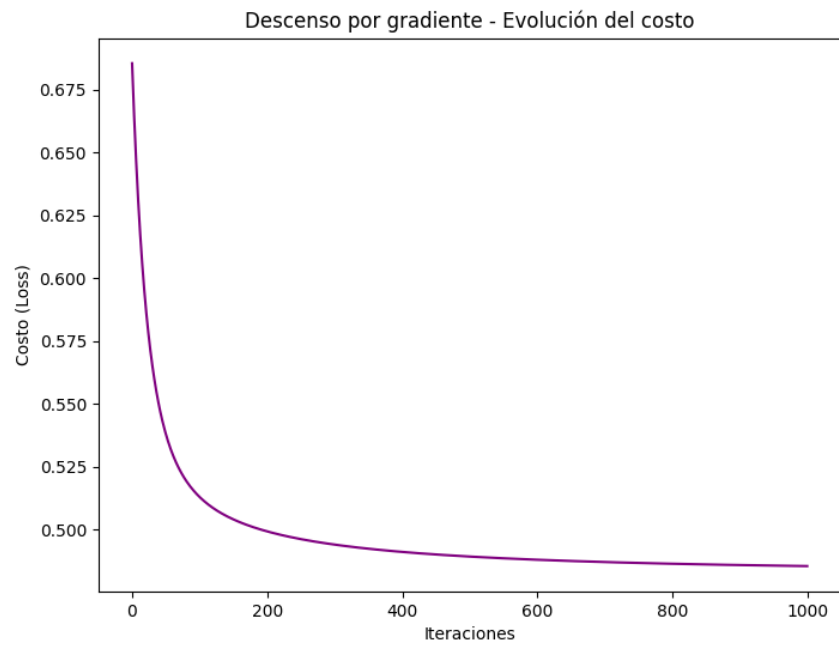


Figura 1: Curva de pérdida durante el descenso por gradiente.

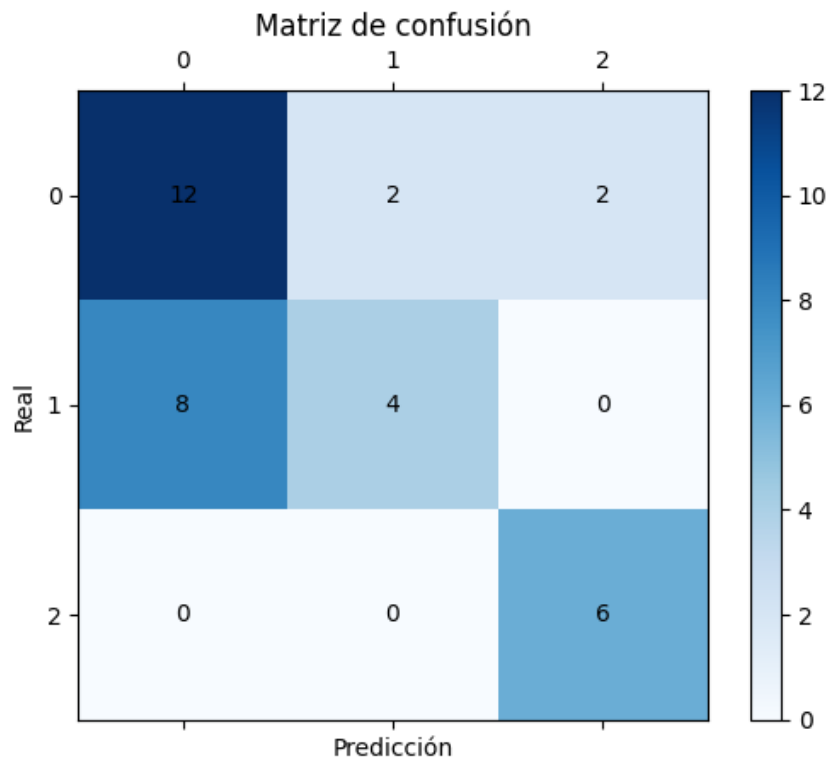


Figura 2: Matriz de confusión del modelo.